Pyrefra user guide

This guide is meant to suggest a work flow for refraction seismic data treatment with pyrefra and to give some hints for options which are perhaps not evident from the user manual. So it is thought to have the inverse approach of the manual, not describing the different options, but the methodological approach.

In several places, other python programs than pyrefra are mentioned. They are located in folder Utility_programs. Explanations for each program are found in the comment block (“docstring”) at the top of the file.

A Acquisition and file preparation

1) During the **acquisition** pay attention to always indicate correctly the shot point number and the receiver points used. This depends evidently on the acquisition system you are using.

2) The **file names** must contain just before the extension a number that is different for each recorded shot. If your acquisition system does not do this automatically, you must modify the names. Something like “first_shot1.seg2”, “second_shot1.seg2” etc will not work, since the program will find the number 1 for all files and crashes. Rename to “first_shot**1**.seg2”, “second_shot**2**.seg2” etc. The numbers are not necessarily contiguous. You may use program **seg2_renumber.py** for this.

3) Since many acquisition systems do not store precise shot and receiver locations in the headers, write files **receiver.geo** and **shots.geo**. The numbers in the first row of these files must coincide with the shot point and receiver point numbers stored in the header. If after starting PyRefra it crashes while reading the files, the problem is often due to **errors in the storage of point numbers** (no information at all or always the same number for shots and receivers. To check the contents of the file and trace headers of seg2 files, you may use program **header_info.py**. For segy files, use Seismic Unix “**surange**”.
   The coordinates should be defined in a **local coordinate system** where the profile runs in X or in Y direction. For plotting purposes, only the coordinate is used that has the largest variation. If the positions have been determined by differential GPS, geophone positions are more precise than with a measurement tape, but in general, the line runs in an oblique direction. In this case, use program **turn_coordinates.py** which projects the coordinates onto the X axes (maintaining the variability of the points around a straight line in the Y coordinate) and the first point given in receivers.geo or shots.geo is shifted to the origin.

4) If there are **errors in the shot point numbering** or if **receiver numbers** are not stored in the headers, create “**file_corrections.dat**” to correct (see PyRefra manual). The numbers in the first row of this file correspond to the number of each file found in the file name just before the extension.

B Picking

Picking is the core of the program. Here are some recommendations:

5) Start with zooming on the times of first arrivals. Do this on the shot with the largest offsets. If there are many receivers per shot (say more than 30), it may be good to zoom in a second step also on the offsets, choose, e.g., the first 30 traces. The resolution will be better and it will be easier to place the picks. When you do this, do picking on the visible traces, finish with right click and then use the keyboard arrows to take the next traces to the right or to the left and restart picking of those traces. CTRL-Z goes back to the former zoom. Anyhow, if you change the shot point, the offset zoom is reset such that all traces are visible, whereas the time zoom is maintained.

6) PyRefra offers **manual and (semi)automatic picking**.
For **manual picking** press M. A stippled blue line is plotted indicating the air wave. Pay attention not to pick arrivals along this line if you are not sure that it coincides with the wave passing through the ground. The air wave has usually much higher frequencies than the ground wave. For manual picking, the **plotting amplitudes** may be very important. If the amplitudes are too small, one does not see the first arrivals. If, however the amplification is too strong, they may be clipped such that it becomes impossible to distinguish between signal and noise. Therefore, it may be necessary to pick different zones of a seismogram section with different amplifications. Sometimes, applying AGC instead of trace normalization gives better results.
Concerning **automatic picking**, Amplitude and LTA-STA routines work only acceptably well with very good data. In order to use **Correlation Picking**, it is recommended to pick first one shot manually and then try the other ones with the correlation method, since the program uses picks made at nearby shots as guidance. For a given shot and receiver, it searches picks made at neighbouring shots made at the same offset and searches a local maximum of the correlation function near to this time. In near-surface data, the main problem with correlation picking is that the frequency content often changes notably with offset, increasing the time difference between the onset and the first maximum and due to this, placing the picks not necessarily at the onset.
If automatic picking gives very bad results, use **CRTL-E** to **erase all picks** of the actual shot and restart with other parameters or with manual picking.
Another possibility, which is actually **my preferred method**, is to use existing picks as guidance. For this enter **Manual Picking** (press M). A first shot has to be picked manually. For subsequent shots, in addition to the blue stippled line also a green one is plotted. This line shows the position of picks at nearby shots done at the same offset. Pressing **SHFT + left mouse button** accepts these times for the picks.
If the picks are not too bad but not perfect either, use **CTRL-M** to **shift picks manually** to the wanted position (use up and down keyboard arrows to shift a pick by one pixel; together with SHFT, pixels are shifted by 10 pixels at a time, together with CRTL by 1 ms). You may choose the picks to be shifted with a mouse click or go with the right/left keyboard keys from one trace to the next.
For the inversion, every pick should be assigned an **uncertainty**. Automatic picking sets the uncertainty by default to 2 samples (4 samples if the data have been filtered before (see later)). When doing **manual picking**, the user must do two clicks, the first one at the position of the pick, the second one marks the uncertainty which is supposed to be symmetric around the pick. A double click sets the default uncertainty (2 or 4 samples).

If the **uncertainty should be modified** (usually increased for automatic picking), press **SHFT-U** choose the pick with a mouse click and increase/decrease uncertainty with the up/down arrow of the keyboard. Right/left keyboard arrow may be used to pass from one trace to the next. Picks are written to file **picks.dat** as soon as right click is detected for manual picking or modification of picks (location and uncertainty) or as soon as automatic picking has successfully finished for a shot.

7) It may be necessary to **filter** data in order to increase signal-to-noise ratio. Frequency filter and velocity filter are available. Since filtering modifies the signal form (especially near the shot point), do it carefully. In any case, **do picking on unfiltered data wherever it is possible**. This is also the reason why default uncertainty is increased for filtered data. If you want to do automatic picking after having done already some picks, zoom in on traces not yet picked, since automatic picking is on the one hand effective only for traces shown actually on the screen, but on the other hand defines picks for all of these traces, i.e. if a trace has already been measured, it will give a second pick for this trace.

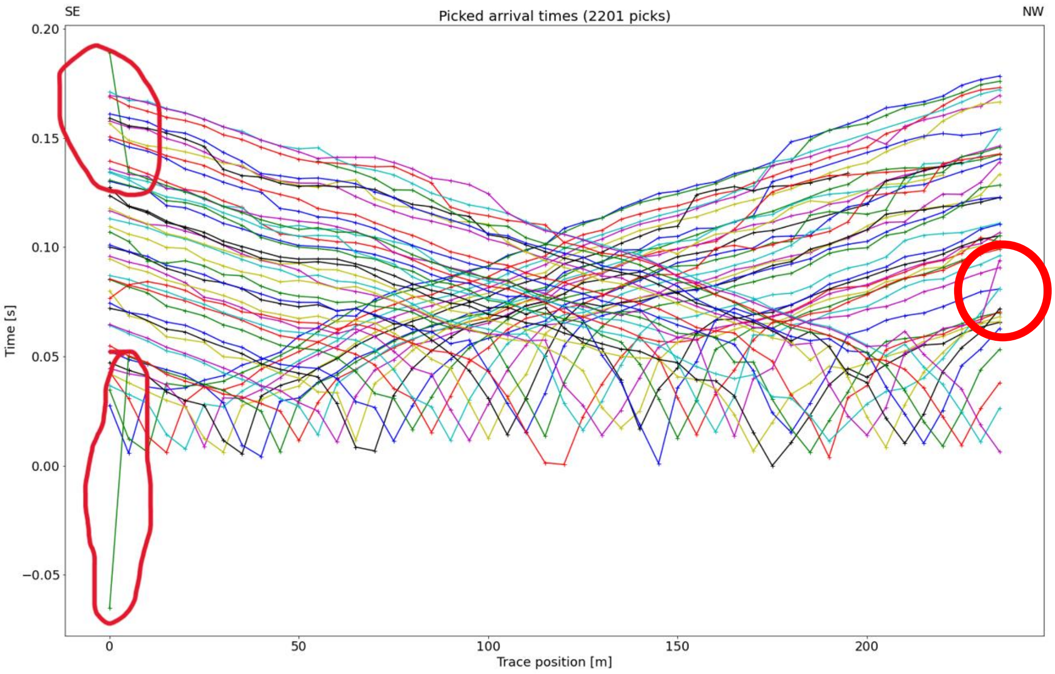
8) Up to 5 picks may be defined for one trace (e.g. P and S), but it is not recommended to use this possibility, since the tomographic inversion will use the average of the times as input data. If you pick **P and S wave arrivals**, do this in two steps and save each one in its own file (rename an existing picks.dat e.g. to picks_P.dat before picking S arrivals and if you want to work again with the P-wave arrivals, rename picks_P.dat back to picks.dat. The program does not do joint inversion of P- and S-wave arrivals.

C Quality control of picks

Mostly, refraction seismic data are easy to pick near the shot point but become more and more noisy with distance. Therefore, although one does the very best to measure correct picks, it may be that, e.g., a wrong phase is picked or that a noise amplitude is mistaken for the real onset. Before trying to do a tomographic inversion, it is therefore important to check the picks. The program offers a bunch of possibilities for quality control:

9) A first check which should be done is to plot the data in **Distance Gathers** (press the letter "d" for this, or go to "Display -> Distance Gather"). There, the traces are plotted as function of the midpoint between shot and receiver. If you have, e.g. a shot at 40m and a receiver at 50m, the trace will be plotted at 45m. But also if the receiver is at 40m and the shot is at 50m, the trace will be plotted at 45. So, you will have two traces at the same position and the ray paths should be exactly the same and therefore also the picked times should be exactly the same, which is often not the case. If they are not the same, they should be corrected (Move Picks, CTRL-M).

10) Plot all measured picks (**CTRL-P or Picking -> Plot all Picks**). This gives a plot like this:



where clear measurement errors will be immediately visible.

11) Also the **pseudo-velocity plots** (V or Utilities -> Pseudo-velocities) may highlight errors through local anomalies

12) **After having run a tomographic inversion**, analysis of the **results** may reveal problems. On the one hand, in **image G**, are shown the average misfits of shots and receivers. It may be worthwhile to check the shots or the receivers (use Receiver Gather, R) for which the misfits are largest. There may be a problem of trigger time or receiver positioning. In **image D**, the travel times are shown where the color scale may reveal shots (horizontal lines) or receivers (vertical lines) for which the times are systematically smaller or larger than those of neighboring ones. **Image E** shows misfits for all measured shot-receiver combinations. In the ideal case, the distribution of misfits should be random, but it may happen that there are systematic differences (large ones for a single shot or receiver, positive differences for positive offsets and vice versa…). Also in this case, go back to the data and check your picks. To check measured vs calculated picks and see where the main problems are located, a good means is to use the option **Picks -> Plot_calculated_times** on the main screen. This will show the calculated travel times as a continuous green line. Often, the first arrivals are not very clear and it may be that for one shot, one decides to pick one phase and for other shots, one picks a different phase. In the case of incompatible picks, the inversion result will do a kind of averaging. If therefore the picks of one shot are far from the calculated results, this usually means that there is a problem with this shot, since most nearby shots were picked differently and may be a different phase, nearer to the calculated line should be picked.

D Tomography

The tomography module is based on pyGIMLi. Pyrefra does not use the full capabilities of pyGIMLi. E.g., only the default gridding is possible. However, a few important parameters may be defined by the user:

13) The **thickness of the model** (maximum depth): by default, the thickness is proposed to be 1/3 of the maximum offset. If after the inversion, you see that the rays don't reach that depth, you may restart the inversion with a smaller thickness. If, on the other hand, the rays reach the maximum depth over a relatively long distance, this means that the thickness was not large enough. Restart with a thicker model.

14) **Starting model**: Pyrefra allows for the moment only a very simple starting model: the velocities at the top and the bottom of the model are fixed and between them a constant gradient is applied. The user may thus run the inversion with different top and bottom velocities, which may give different results if the velocities are not well determined by the measured travel times.

15) **Smoothing parameters**: Since the initial model is usually far from the optimum model, the matrix to be inverted may be near to singular and therefore, it should be regularized. One possibility, used here, is to add a constant to the diagonal which effectively results in smoothing the final model. The inversion is done iteratively and one may suppose that with every iteration, the model approaches an optimum and the inversion becomes more and more stable. Therefore, the smoothing factor may be decreased at each iteration step. The factor by which the smoothing is multiplied may be defined by the user. Finally, often seismic velocities vary mainly vertically and less horizontally. If one wants to simulate this type of model, it is possible to reduce vertical with respect to horizontal smoothing by setting Z-smoothing to a value smaller than one. All these smoothing parameters may be changed and different tomographic inversions should be run with different parameters to analyse the stability of the resulting model.

16) **Search range**: pyGIMLi allows limiting the search range for velocities, i.e. to give minimum and maximum allowed velocities. In practice, I do not completely understand this procedure in pyGIMLy. My experience is that if no limits are given at all, the resulting velocities are really unrealistic. However, as soon as limits are given, the inversion behaves well, independent of the value given by the user.

17) **Model uncertainty**: Travel time tomography does not give unique results. On the one hand, the reliability of a model depends on the ray coverage. A measure of this ray coverage is shown in image C of the inversion result. The higher the coverage, the better resolved is the model. This, however is only a qualitative measure. Since the final model depends on the above defined parameters, another way to get a more quantitative measure of uncertainty is to do many runs with different starting models and different smoothing parameters (as mentioned, the velocity limits, as soon as they are given, seem to have very little influence). One may then calculate the variance of the results for different areas of the model. This is not necessarily straight forward, since the used grid may vary from one run to the next. A possibility is to regrid the velocities of every model onto the same regular grid and calculate the variance for every grid point. This will not be a mathematically sound uncertainty, but it gives an idea.

18) Program **Extract_velocities.py** allows extracting the depth of a velocity isoline (shallowest occurrence of a given velocity along the line), a vertical profile at a given coordinate X of a horizontal profile at a given coordinate y.

19) If needed, the velocity model may be re-projected onto GPS coordinates, e.g. to integrate the model into a GIS system. Program **Tomo_coordinate_projection.py** does this job.